

# Mouse Phenome Database (MPD)

Terry P. Maddatu, Stephen C. Grubb, Carol J. Bult and Molly A. Bogue\*

The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

Received September 26, 2011; Revised October 25, 2011; Accepted October 26, 2011

## ABSTRACT

**The Mouse Phenome Project was launched a decade ago to complement mouse genome sequencing efforts by promoting new phenotyping initiatives under standardized conditions and collecting the data in a central public database, the Mouse Phenome Database (MPD; <http://phenome.jax.org>). MPD houses a wealth of strain characteristics data to facilitate the use of the laboratory mouse in translational research for human health and disease, helping alleviate problems involving experimentation in humans that cannot be done practically or ethically. Data sets are voluntarily contributed by researchers from a variety of institutions and settings, or in some cases, retrieved by MPD staff from public sources. MPD maintains a growing collection of standardized reference data that assists investigators in selecting mouse strains for research applications; houses treatment/control data for drug studies and other interventions; offers a standardized platform for discovering genotype–phenotype relationships; and provides tools for hypothesis testing. MPD improvements and updates since our last NAR report are presented, including the addition of new tools and features to facilitate navigation and data mining as well as the acquisition of new data (phenotypic, genotypic and gene expression).**

## INTRODUCTION

Phenomics, the comprehensive study of phenotypes, complements advances in knowledge from genomics, pharmacogenomics and other -omics approaches. Most human diseases are complex, polygenic and subject to modifications by environmental, genetic and epigenetic factors. A phenomics approach requires collecting comprehensive high-dimensional phenotypic data, integrating it with other types of diverse biological data, and evaluating it with computational tools and bioinformatic approaches.

MPD is the only database of its kind—amassing, annotating, integrating and maintaining quantitative phenotypic data for the laboratory mouse, and providing online tools to access and analyze those data. MPD is an established community resource and research tool enabling investigators to access thousands of data points from hundreds of strains, saving time and money by precluding the need to (re)characterize strains. MPD provides many tools to facilitate mouse research. For example, MPD enables users to:

- Compare strains to identify mouse models of human disease.
- Select optimal genetic backgrounds for new or existing mutations.
- Take advantage of the MPD Protocol Library containing detailed, illustrated procedures that have been validated by experts in their fields (and further, to view data generated from these protocols to confirm experimental results).
- Discover phenotype–genotype relationships and
- Validate and refine results from traditional mapping studies.

More powerful mapping tools are increasingly essential and indispensable in the analysis of complex traits and systems genomics to dramatically reduce the broad, unwieldy intervals of previously identified quantitative trait loci (QTL) containing hundreds (if not thousands) of genes, down to the gene level. For several years now, the community has had access to high-density SNP data for 16 inbred strains at over 8 million genome-wide locations through the NIEHS-Perlegen resequencing effort (1) and for 94 strains at 132K locations provided by the Mouse HapMap Project (2). These data were used to develop a novel high-density genotyping array, the Mouse Diversity Genotyping Array (3) (see also <http://cgd/tools/diversityarray.shtml>). These SNP data sets are available in MPD: Perlegen2 (4), Broad2 (5) and CGD2 (6), respectively. MPD has Mouse Diversity Array data for 123 inbred strains at 550K locations. In addition, the Wellcome Trust Sanger Institute recently released whole-genome sequence data for 17 inbred mouse strains and is generating transcriptome data (RNA-Seq) for these

\*To whom correspondence should be addressed. Tel: +1 207 288 6016; Fax: +1 207 288 6079; Email: molly.bogue@jax.org

strains (7,8). Taken together with high-quality phenotype data, transcriptome and dense genotypic data enable the identification of regions of the genome that may be causal to a phenotype of interest through correlative analyses. Sanger SNPs, indels and structural variants will be added to MPD and made available in the coming months.

MPD history, standards, schema and other topics have been presented elsewhere (9–11). Here, we review a few salient points regarding MPD contents. The recommended basic study design is a strain survey format, e.g. 10 females and 10 males of a large number of genetically stable, reproducible strains (such as inbred strains). Acquired data sets are annotated and formatted to meet a set of defined MPD standards. Each data set is accompanied by a detailed protocol; health status and environmental parameters of the test animals; and any other information essential to understanding and evaluating the data. Data and documentation are then bundled together into an MPD ‘project’, which is given a descriptive title and annotated with submitter information, references and funding acknowledgements.

MPD phenotypic data sets are typically quantitative, although we do house categorical data such as MHC haplotypes and a growing collection of annotated images. Data from microarray gene expression studies are considered a special class of phenotype data and handled accordingly. Summary statistics are computed as part of the MPD accessioning process, where the unit of analysis is an MPD ‘measurement’ with strain (by sex) being the analysis group. Each measurement in a data set is classified and integrated in the MPD phenotype category structure. Male and female data are not combined. Data from different MPD measurements are not averaged together although these data may be displayed in the same plot for comparison (composite/consensus view). As part of the accessioning process, regression analysis is performed on all measurements (pair-wise) to identify compelling correlations, which may indicate common genetic determinants or overlapping biological pathways. Results are stored to optimize and support queries based on correlations. Strain genotypes are collected and stored in MPD so that phenotypic and genotypic data can be linked or visually juxtaposed, facilitating the ability to determine how allele-specific variations translate to variation in mouse phenotype.

Our last NAR update was published in 2009 (11). Here, we summarize new data and data types, and describe major improvements in MPD functionality. This update refers to [Supplementary Data](#).

## CONTENTS

‘MPD current contents’ are summarized in [Table 1](#), and a snapshot of ‘What’s New’ is shown in [Table 2](#). Users can search the database by our keyword search, or browse by strain, phenotype category, data submitter, interventions, study design and other convenient groupings ([Supplementary Data](#), Navigation and Finding Data: S1–S8). Briefly, over 750 strains of mice are represented

**Table 1.** MPD current contents

Phenotype measurements (including pending) [S7] collectively ~600 strains	~3000
SNPs [S11]	
Total calls (excluding imputed)	~350 million
Genomic locations	8+ million
Number of strains with data	450+
Gene expression strain surveys [S12]	8
Number of probesets	280K
Strains [S2, S3]	761
Strain types	
Inbred	208
Recombinant inbred strains	330
F1 hybrids	87
Wild-derived inbred	70
Chromosome substitution strains	38
Other	28
Recombinant inbred panels	9
AKXD	
AXB	
BXA	
BXD	
BXH	
CXB	
ILSXISS	
LGXSM	
SWXJ	
Chromosome substitution panels	2
C57BL/6J-Chr <sup>A/J</sup> /NaJ	23
C57BL/6J-Chr <sup>PWD/F1</sup> /ForeJ	19
Animal vendors and repositories	10
Animal Resources Centre (AUS)	
B&K Universal (UK)	
Charles River	
Harlan Laboratories	
The Jackson Laboratory	
Nihon Clea, Tokyo (Japan)	
Pasteur Institute (France)	
RIKEN BioResource Center (Japan)	
National Institute of Genetics—Shigen (Japan)	
Taconic	

in MPD with phenotypic and/or genotypic data available. Strain names are kept updated with current nomenclature, and linkouts to other databases are provided. [Table 1](#) provides an inventory of strains, strain panels and strain sources (also Search or Browse Strains: S1, S2, S3). Investigators associated with MPD as data contributors or as formal collaborators represent approximately 150 institutions from 14 countries supported by over 80 funding agencies and research foundations worldwide (Investigators: S4; Large-Scale Initiatives: S5).

## Phenotype data

MPD contains approximately 3000 phenotypic strain survey measurements, collectively representing approximately 600 of the 750+ strains in MPD (Find Measurements: S6, S7, S8). Most measurements (70%) include data for both sexes. On an average, 20 strains are tested per measurement, but as many as 72 strains have been tested in a single project (CGDpheno1, pending). Currently available phenotypic measurements are classified as baseline (64%), aging-related data (14%) or controlled studies of intervention effects

**Table 2.** Summary of What's New since our last NAR update

MPD categories with new data	Treatment and intervention studies [S8] (with new data)
Appearance	Dietary interventions
Coat color	Capsaicin
Behavior	Diet restriction
Anxiety	High-fat diet
Exploratory	High-fat diet + ethanol
Learning and memory	Sucrose
Locomotor activity	Drugs and alcohol
Motor coordination and activity	Acetaminophen
Social	Ethanol
Stress reactivity	Fluoxetine
Wildness	Haloperidol
Blood	Isoniazid
Clinical chemistry	Lithium
Hematology	$\beta$ -adrenergic drugs
Lipids	Environmental insult
Body composition	Trichloroethylene
BMI	Infectious diseases
Fat pads	Candida
Body metrics	Influenza A ( $H_1N_1$ )
Size, weight, growth curve	Malaria
Bone	Lifestyle and healthspan
Mineral density and content	Aging
Metrics	Exercise
Brain	
Drug concentration	
Gene expression	
Cardiovascular	
Echocardiography	
Blood pressure, heart rate	
Cell and tissue damage	
Apoptosis	
Chromosome instability	
Ear	
Balance	
Hearing	
Endocrine	
Adrenal morphology	
Hormones	
Immune system	
Cell counts	
Immunoglobulins	
Response to infection	
Kidney	
Urinalysis	
Vesico-ureteral reflux	
Liver	
Pathology	
Homeostasis	
Gene expression	
Longevity	
Lifespan	
Metabolism	
Body temperature	
Energy	
Food and water intake	
Nervous system	
Autonomic function	
Neuron projection morphology	
Neurogenesis	
Neuromuscular function	
Sensorimotor function	
Neurosensory	
Nociception	
Sensory gating	
Organ metrics	
Eye	
Kidney	
Liver	
Spleen	
Reproduction	
Fecundity	
Female	
Male	
Sex distribution	
Respiratory	
Airway resistance	
Breathing pattern	
Lung compliance	
Lung volume	

(22%), e.g. administering drugs or high-fat diet (Interventions: S8). MPD has acquired and formatted hundreds of experimental procedures, housing a library of protocols where detailed procedures and assays are maintained so that the research community may benefit from their use (Protocols: S9, S10).

## Genotype data

The current MPD SNP database contains ~350 million SNP calls (count excludes imputed SNPs) obtained from approximately 30 publicly available or contributed SNP sources, including Broad, Celera, Center for Genome Dynamics, Merck, Perlegen (NIEHS), Wellcome Trust, Novartis (GNF) and The Jackson Laboratory (MPD SNP Database: S11). MPD makes SNPs available in sets for RI panels (8 total), C57BL/6 substrains (13 variants) and 129 substrains (7 variants). In total, MPD has SNP data for >450 verified strains (Find Strains with SNPs: S2, S3). MPD combines SNP data with gene annotations from external resources such as Mouse Genome Informatics (MGI) (12), NCBI dbSNP (13) and Ensembl (14).

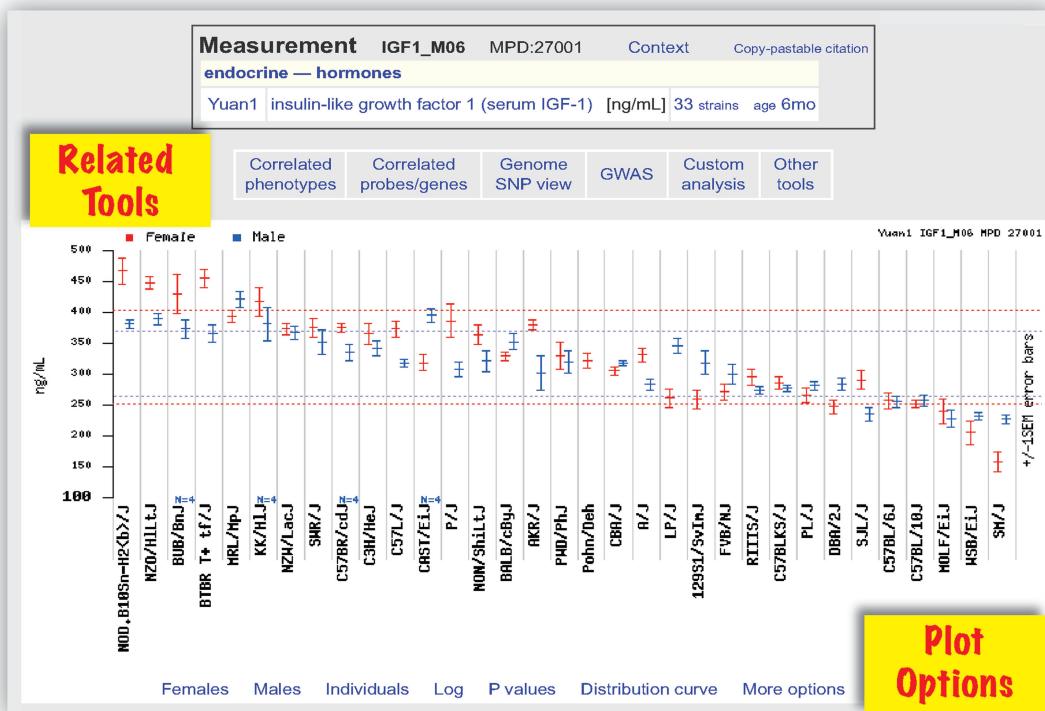
## Gene expression data

In addition to new phenotypic and genotypic data, MPD has recently begun accessioning gene expression strain survey data. Currently, there are eight gene expression projects, representing ~280K probesets. Four of these projects are part of a treatment/control paradigm, and phenotype data are available in parallel. These intervention studies are from two laboratories studying the effects of a high-fat diet [Shockley1 (15,16)] and an environmental insult (trichloroethylene exposure) [Rusyn2 (17,18)]. Thus far, data are from either whole brain or liver; efforts are underway to obtain data from other tissues and cell types. Most data sets are inbred strain surveys (10–30 strains), and another is for the ILSXISS recombinant inbred strain panel (61 strains) [GX-Tabakoff2 (19)]. Functionality of the new gene expression interface is showcased below in detail (see also Gene Expression: S12).

Since our last reporting, we have accessioned over 70 new phenotype or gene expression projects from over 100 investigators representing >40 institutions. See Table 2 for ‘What’s New’ in MPD categories and intervention groups. Several of these new projects are important drug studies, including acetaminophen toxicity [Threadgill1, Threadgill2 (20–23)], haloperidol effects on behavior [Crowley1 (24,25)] and effects of  $\beta$ -adrenergic drugs on heart function [Maurer1 (26,27)] (Projects: S6; Interventions: S8). Most have associated publications, providing assurance that the study has been peer-reviewed and the data validated by experts.

## FUNCTIONALITY

MPD provides many options to view, analyze and download data (MPD Toolbox Demo: S13). Since our last NAR update, we have made significant improvements



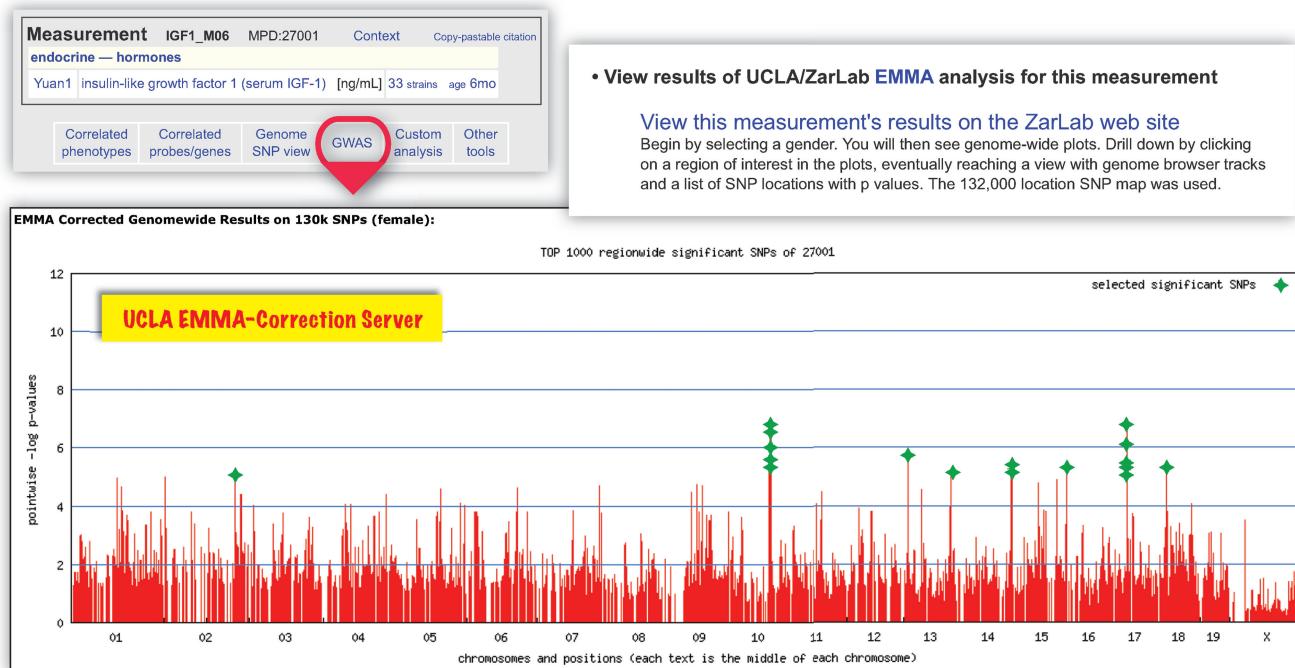
**Figure 1.** Phenotype data: new layout, viewing options and tools. An MPD summary plot is shown from an aging study measuring IGF-1 levels in males and females of 33 inbred strains at 6 months of age [Yuan1 (28); see also accompanying paper (29)]. From this plot, users have several choices for viewing (see plot options), and there are one-click options for deploying relevant phenotype measurements that correlate significantly, finding genes with correlated expression patterns and performing EMMA-corrected genome-wide association analysis (GWAS). For most MPD measurements with at least 20 inbred strains tested, the GWAS option is a direct link to pre-computed results hosted on the UCLA EMMA website (see text). About the plot: strain means are shown ( $\pm 1$  SEM); dotted horizontal lines indicate 1 SD for the overall strain mean (red = female; blue = male). All words in blue are links to tools, other views or additional information. For a continuation of this example, see Figure 2 (GWAS) and Figure 3 (gene expression).

and added new or expanded functionality to the MPD Toolbox. Here, we present a comprehensive example to demonstrate the power and utility of MPD and to highlight some of our new tools and features. We start with MPD phenotype data (IGF-1 levels at 6 months of age) and use new MPD tools to confirm and further narrow published mapping results (Yuan1) (28,29). This example spans Figures 1–3 and illustrates the following new or updated features:

- Improvement of various tools with new layouts, more options and links to other tools and information (Figure 1).
- Development of interfaces to facilitate association mapping using the UCLA EMMA-Correction Server (see below and Figure 2).
- Implementation of a user interface for working with gene expression data and providing tools for finding significant correlations to phenotype data using an innovative variance-based, data-screening feature to minimize false positives (Gene Expression: S12; Correlations Center S14) (Figure 3) and
- Implementation of a ‘workbook’ feature to help users pare-down and gather candidate genes with an option to generate an output of gene IDs for quick

copy–paste into other bioinformatics resources (Figure 3).

MPD has been exploring and evaluating methods for genome-wide association studies (GWAS) that may be developed into general tools for MPD users. One approach that has received a lot of attention is the ‘efficient mixed-model method’ (EMMA), which corrects for inbred strain population structure thereby significantly reducing the number of false positive associations (30). GWAS results for selected MPD measurements (e.g. testing at least 20 strains) are available in collaboration with the UCLA Computer Science and Bioinformatics group (ZarLab) using EMMA and the Mouse HapMap for 94 inbred strains at 132K locations (2). Wild-derived strains are automatically excluded from these analyses (see more details in Figure 2). For ineligible MPD measurements (<20 strains), MPD provides an option to produce input-ready files specifically formatted for a real-time run on the UCLA EMMA Server (mouse.cs.ucla.edu/emmaserver). Users can choose to include wild-derived strains and are able to further manipulate the input file, e.g. to delete specific strains, and to select other SNP data sets for interrogation.



**Figure 2.** New GWAS tools. Continuing with the IGF-1 example [Yuan1 (28)], we performed a genome-wide association analysis by clicking on the ‘GWAS’ option above the plot shown in Figure 1 (the relevant part of this panel is shown here for convenience). We found significant EMMA-corrected peaks for females on chromosomes 2, 10, 13, 14, 15, 17 and 18 (lower panel; green stars were manually overlaid on this plot to ensure visibility of peaks). Interestingly Leduc *et al.* (29) (the data submitter’s publication for Yuan1) found three of the same *in silico* QTLs using another method, haplotype association mapping (HAM), which employed the Hidden Markov Model and a set of SNPs at 70K locations (includes imputed calls). We chose to concentrate on the most significant peak identified by these methods on Chr 10 in the vicinity of the *Igf1* gene. Using other supporting data sets and bioinformatics tools, Leduc *et al.* narrowed their results from over 45 genes annotated in the QTL interval to 21 candidate genes. We wondered if we might narrow these results even further by using another new MPD tool for finding significant gene expression correlations (see Figure 3). The new GWAS tool is provided in collaboration with the UCLA Computer Science and Bioinformatics group (ZarLab) (30). MPD is not responsible for the UCLA website or EMMA Server ([mouse.cs.ucla.edu/emmaserver](http://mouse.cs.ucla.edu/emmaserver)).

In addition to the highlighted examples (Figures 1–3), MPD has also implemented the following since our last NAR update:

- Correlations Center: centralizes, organizes and explains all MPD correlation operations on a single webpage; and provides convenient links to relevant tools (Correlations Center: S14).
- Private Analysis Portal: users enter their own experimental data and run real-time regression analysis across the entire MPD to identify statistically significant correlations to existing phenotype and gene expression data (Private Analysis Portal: S15) and
- Tool to generate ratio or difference vectors for user-selected MPD measurements: users can view plots and identify independent MPD measurements that are highly correlated to the *de novo* generated vector (MPD Toolbox Demo: S13; Correlations Center: S14).

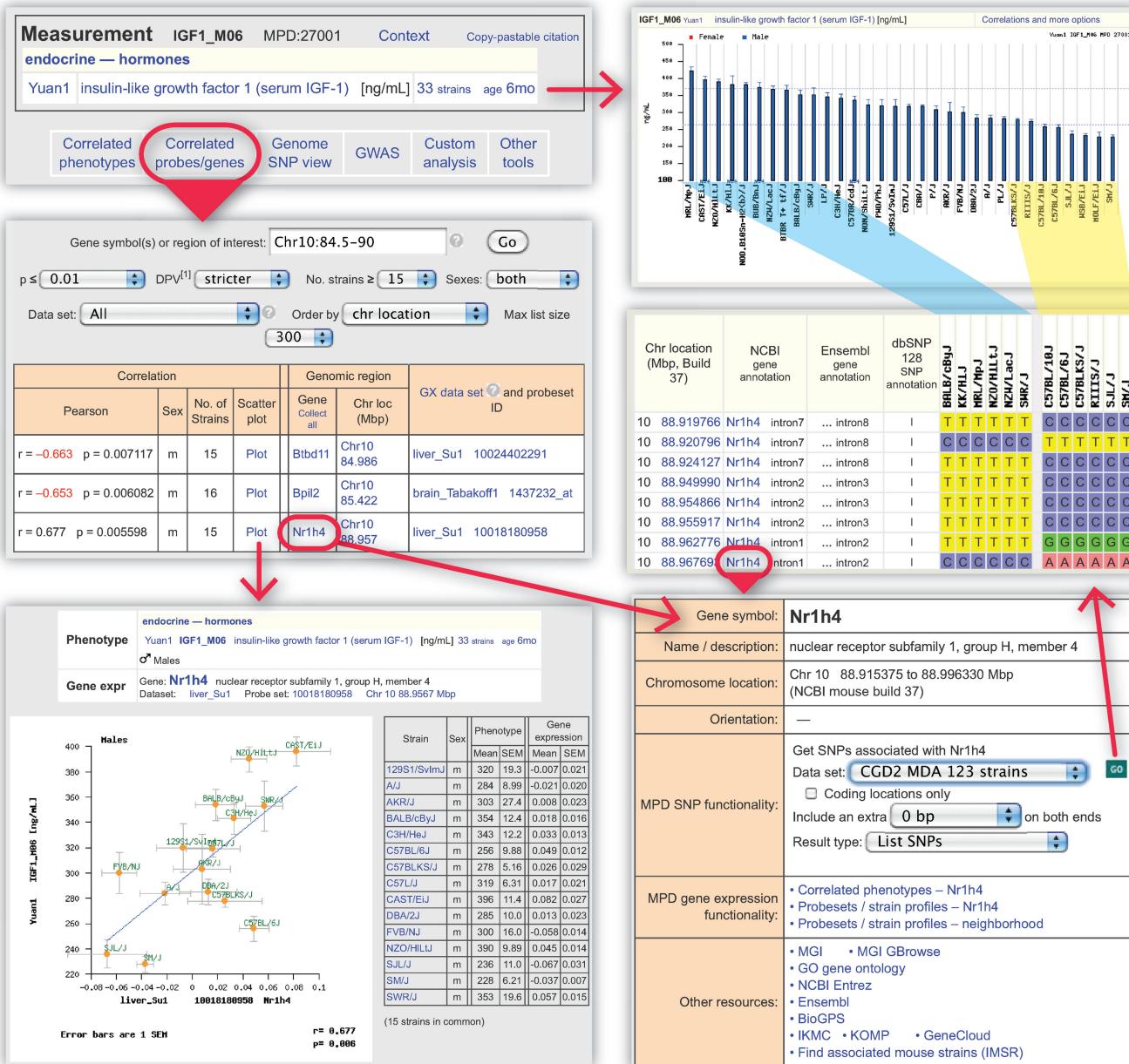
#### Additional supplementary data

Individual animal data, summary statistics and other metadata are available for downloading in bulk or in customized data sets specified by the user (Download

Center: S16; SNP Downloads: S11). For convenience, we are providing [Supplemental Material](#) for general content and information (MPD Content and Quick Facts: S17) and a user’s manual (MPD User Manual and FAQ: S18).

#### HIGH-LEVEL OVERVIEW OF IMPLEMENTATION

The specifics of our implementation have not changed significantly since our previous update (11). MPD runs on a SUSE Linux/Apache 2 system using open source database, HTML generation and graphics generation software, with custom-written C modules used in performance-critical areas. Phenotype data sets are typically submitted by data contributors as Excel spreadsheets. Genotype and gene expression data sets are usually downloaded by us from open public resources. *Ad hoc* processing methods are used to integrate these data into our standardized framework. In conjunction with data contributors, MPD staff creates supporting documentation in HTML. We add content and perform testing and QA on an internal development node; then this instance is copied over for public release. MPD follows an agile development philosophy with rapid prototyping and frequent updates; version numbering is not used. For



**Figure 3.** New gene expression interface. Continuing with the IGF-1 example [Yuan1 (28)] from Figures 1 and 2, we interrogated the Chr 10 QTL interval (84.5–90.0 Mbp) identified by both association mapping methods (HAM and EMMA) using the new MPD gene expression interface. Users can simply click on the *correlated probes/genes* option above the plot shown in Figure 1 (the relevant part of this panel is shown in the upper-left panel for convenience), or users may choose to begin in the Correlations Center (S14). Of those 21 candidate genes identified by Leduc *et al.* (29) (see Figure 2), we went on to confirm that 6 of those genes had compelling expression profiles consistent with IGF-1 levels ( $P < 0.05$ ). Three of these genes are listed in the middle-left panel (see Note 1 for more information about this panel). As indicated, three probesets from two gene expression projects [GX-Su1 (31,32), GX-Tabakoff1 (33)] were identified using various filtering and sorting options that are available with this tool. Users can adjust stringency to broaden or narrow results by selecting  $P$ -value cutoff and setting the *data point variance* (DPV) quality filter where stricter settings will omit probesets that have larger standard errors relative to the range of strain means. Users may further filter results by minimum number of strains, sex, specific data set and chromosome. Results can be sorted by  $P$ -value, correlation coefficient, location order, MPD project or MPD category. A phenotype-gene expression scatterplot is available through the ‘Plot’ link. IGF1\_M06 versus *Nr1h4* for males is shown in the lower-left panel. By clicking on gene name, we accessed the gene directory page (lower-right panel, and see Note 2 below), to deploy the 550K SNP panel [CGD2 (3,6)] specifically for the gene of interest to compare the segregation of phenotype with polymorphisms. Using MPD SNP tools, we confirmed that three of the six genes have polymorphic intronic or UTR regions, consistent with the notion that these genes are differentially regulated. The corresponding SNP results for *Nr1h4* are shown (middle-right panel) where strains are grouped according to ‘hi’ (left) or ‘lo’ (right) IGF-1 levels. These strains fall into two major haplotypes (wild-derived strains fall into a third haplotype, not shown). For ease of comparison, an optional view of the male IGF-1 data is shown in the upper-right panel [color-guides from the SNP results to plot: blue = high IGF-1 levels, yellow = low IGF-1 levels (see also, strain distribution in Figure 1)]. All words in blue are links to tools, other views or additional information. While we demonstrate that MPD can help identify (or further verify) candidate genes exhibiting differential expression, we also note that caution should be used in making claims of causality based solely on microarray data, SNPs and functional relevance of positional candidates. All MPD phenotype-gene expression pair-wise correlations are available for bulk download (Download Center: S16). Note 1: The ‘collect gene’ option in the middle of the gene expression

(continued)

more documentation, see MPD ‘BioDBcore Statement’ (S19). Information for developers is also available (For Developers: S20).

## FUTURE MPD DIRECTIONS AND AN INVITATION TO INVESTIGATORS

MPD will continue implementing useful linkages to other databases and populating the database with phenotype, genotype and gene expression data. We continue to refine MPD data classifications with relevant ontologies, optimizing search and navigation functionality as well as facilitating data integration within MPD and across databases. We are in the process of loading Sanger SNPs, indels and structural variants derived from whole-genome sequencing of 17 inbred mouse strains (7,8), and we will take advantage of the availability of the RNA-seq data being released by Sanger on this set of strains. RNA-seq data will complement existing MPD data and will enable queries to determine expression levels of different alleles of a gene and to detect post-transcriptional mutations or other genetic alterations.

Many phenotypic domains are currently represented in MPD; however, the acquisition of new data is open-ended with the goal of collecting data on a broader scope (and in some cases to a deeper level for more granular phenotypes or endophenotypes) as well as collecting data generated from new, more sophisticated phenotyping technologies. Although we have added a significant number of phenotypes since our last update (Table 2), we will continue working toward accessioning more data for inbred strains and other strain types as well as for new panels such as the Collaborative Cross (34–36). We would like to take this opportunity to make an appeal for clinically relevant data in toxicology and specific disease areas: cancer, digestive diseases, kidney disease, neurological disorders, drug addiction and infectious diseases. To expand the scope and maximize the utility of MPD, members of the global scientific community are invited to contribute their strain survey data or join us in a coordinated effort to seek funding that will support these studies. This spirit of collaboration has shaped and will continue to guide the future growth and development of MPD. Researchers interested in contributing data or in collaborating on new phenotyping projects should contact us at [phenome@jax.org](mailto:phenome@jax.org). Data submission guidelines are accessible through the MPD homepage (Contributing Data: S21).

**Figure 3.** Continued

results (middle-left panel) is one place where users can take advantage of our new ‘workbook’ feature for paring down and gathering candidate genes. Lists are provided a new window, and checkboxes are supplied to exclude (or include) genes in an updated list. At any time, users may opt to generate a separate clean list of gene symbols or various accession ID options to copy/paste into batch query forms of other bioinformatics resources. Note 2: The gene directory box shown in the bottom-right panel is a hub for information and linkouts to other resources (additional features are added as needed). We provide one-click deployment to SNP tools, finding MPD phenotype measurements that correlate, and information about available gene expression probesets, including tools to compare probesets between gene expression projects and to compare female and male for a single probeset within a project. Users will arrive at a gene directory page from any gene link on the MPD website, from gene searches and from general searches that turn up genes. For Developers wanting to link to our gene directory page, use this URL template: <http://phenome.jax.org/db/qp?rtn=markers/details&reqsym=Abcd1>, where Abcd1 is gene symbol (For Developers: S22).

## CITING MPD AND USER SUPPORT

For general citation of MPD, this NAR article may be used. In addition, the following citation format is preferred when referring to MPD projects or using MPD data sets.

Investigator(s) name (year project posted) Project title. MPD project symbol (e.g. Smith1) and/or accession number (MPD:XXX). Mouse Phenome Database Website, The Jackson Laboratory, Bar Harbor, ME, USA. (URL: <http://phenome.jax.org>), date of download or access.

For more information visit our website and search on ‘citing’.

MPD provides user support through online documentation (S17, S18) and via email ([phenome@jax.org](mailto:phenome@jax.org)). We welcome user input and suggestions, which can be submitted anonymously through our Suggestion Box links at the bottom of most MPD webpages.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1.

## ACKNOWLEDGEMENTS

We thank participating investigators for contributing their data for worldwide access (S4, S5). GWAS-EMMA analyses are provided through a collaboration with the UCLA ZarLab (Eleazar Eskin); in particular, we thank Emrah Kostem for his diligence and efforts to make this possible. We appreciate the critical review of this manuscript by Mary Dolan, Susan M. Bello and Alicia Valenzuela.

## FUNDING

The Jackson Laboratory and National Institutes of Health (grant numbers HG003057, HL66611, AG025707, AG038070, MH071984, DA028420). MPD houses and maintains data from research projects supported by over 80 funding agencies and research foundations worldwide (Funding: S22). Funding for open access charge: National Institutes of Health (DA028420).

*Conflict of interest statement.* None declared.

## REFERENCES

- Frazer,K.A., Eskin,E., Kang,H.M., Bogue,M.A., Hinds,D.A., Beilharz,E.J., Gupta,R.V., Montgomery,J., Morenzoni,M.M., Nilsen,G.B. *et al.* (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*, **448**, 1050–1053.
- Kirby,A., Kang,H.M., Wade,C.M., Cotsapas,C., Kostem,E., Han,B., Furlotte,N., Kang,E.Y., Rivas,M., Bogue,M.A. *et al.* (2010) Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics*, **185**, 1081–1095.
- Yang,H., Ding,Y., Hutchins,L.N., Szatkiewicz,J., Bell,T.A., Paigen,B.J., Gruber,J.H., de Villena,F.P. and Churchill,G.A. (2009) A customized and versatile high-density genotyping array for the mouse. *Nat. Methods*, **6**, 663–666.
- Cox,D. and Frazer,K.A. SNP data, 8.2+ million locations for 16 inbred strains of mice. MPD:Perlegen2. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Daly,M. and Wade,C.M. SNP data in support of a haplotype map, 132,000+ locations for 90 inbred strains of mice. MPD:Broad2. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Center for Genome Dynamics (CGD). SNP data from Mouse Diversity Genotyping Array, 550,000 locations for 123 strains of mice. MPD:CGD2. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Keane,T.M., Goodstadt,L., Danecek,P., White,M.A., Wong,K., Yalcin,B., Heger,A., Agam,A., Slater,G., Goodson,M. *et al.* (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, **477**, 289–294.
- Yalcin,B., Wong,K., Agam,A., Goodson,M., Keane,T.M., Gan,X., Nellaker,C., Goodstadt,L., Nicod,J., Bhomra,A. *et al.* (2011) Sequence-based characterization of structural variation in the mouse genome. *Nature*, **477**, 326–329.
- Paigen,K. and Eppig,J.T. (2000) A mouse phenome project. *Mamm. Genome*, **11**, 715–717.
- Bogue,M. (2003) Mouse Phenome Project: understanding human biology through mouse genetics and genomics. *J. Appl. Physiol.*, **95**, 1335–1337.
- Grubb,S.C., Maddatu,T.P., Bult,C.J. and Bogue,M.A. (2009) Mouse phenome database. *Nucleic Acids Res.*, **37**, D720–D730.
- Blake,J.A., Bult,C.J., Kadin,J.A., Richardson,J.E. and Eppig,J.T. (2011) The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.*, **39**, D842–D848.
- Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Federhen,S. *et al.* (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **39**, D38–D51.
- Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2011) Ensembl 2011. *Nucleic Acids Res.*, **39**, D800–D806.
- Shockley,K., Paigen,B. and Churchill,G.A. Diet effects on blood chemistry and lipids in 10 inbred strains of mice. MPD:Shockley1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Shockley,K.R., Witmer,D., Burgess-Herbert,S.L., Paigen,B. and Churchill,G.A. (2009) Effects of atherogenic diet on hepatic gene expression across mouse strains. *Physiol. Genomics*, **39**, 172–182.
- Rusyn,I. and Threadgill,D.W. Effects of trichloroethylene in males of 17 inbred mouse strains. MPD:Rusyn2. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Bradford,B.U., Lock,E.F., Kosyk,O., Kim,S., Uehara,T., Harbour,D., DeSimone,M., Threadgill,D.W., Tryndyak,V., Pogribny,I.P. *et al.* (2011) Interstrain differences in the liver effects of trichloroethylene in a multistrain panel of inbred mice. *Toxicol. Sci.*, **120**, 206–217.
- Tabakoff,B. and Saba,L. Whole brain gene expression in males of 61 ILSXISS recombinant inbred strains. MPD:GX-Tabakoff2. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Threadgill,D.W., Harrill,A. and Rusyn,I. Drug study: toxicogenetic survey of susceptibility to acetaminophen-induced liver injury in males of 37 inbred strains of mice.
- MPD:Threadgill1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Threadgill,D.W. and Harrill,A. Drug study: acetaminophen pharmacokinetics and drug-induced liver injury in males of 5 inbred strains of mice. MPD:Threadgill2. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Harrill,A.H., Ross,P.K., Gatti,D.M., Threadgill,D.W. and Rusyn,I. (2009) Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. *Toxicol. Sci.*, **110**, 235–243.
- Harrill,A.H., Watkins,P.B., Su,S., Ross,P.K., Harbourt,D.E., Stylianou,I.M., Boorman,G.A., Russo,M.W., Sackler,R.S., Harris,S.C. *et al.* (2009) Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. *Genome Res.*, **19**, 1507–1515.
- Crowley,J.J., McLeod,H.L. and Sullivan,P.F. Haloperidol-induced vacuous chewing movements, extrapyramidal side effects, and spontaneous locomotor activity in males of 27 inbred mouse strains. MPD:Crowley1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Crowley,J.J., Adkins,D.E., Pratt,A.L., Quackenbush,C.R., van den Oord,E.J., Moy,S.S., Wilhelmsen,K.C., Cooper,T.B., Bogue,M.A., McLeod,H.L. *et al.* (2010) Antipsychotic-induced vacuous chewing movements and extrapyramidal side effects are highly heritable in mice. *Pharmacogenomics J.* (doi: 10.1038/tpj.2010.82; epub ahead of print).
- Maurer,F., Beckmann,J., Abriel,H. and Pedrazzini,T. Cardiovascular response to beta-adrenergic blockade or activation in males of 23 inbred strains of mice. MPD:Maurer1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Berthonneche,C., Peter,B., Schupfer,F., Hayoz,P., Katalik,Z., Abriel,H., Pedrazzini,T., Beckmann,J.S., Bergmann,S. and Maurer,F. (2009) Cardiovascular response to beta-adrenergic blockade or activation in 23 inbred mouse strains. *PLoS One*, **4**, e6610.
- Yuan,R., Rosen,C.J. and Beamer,W.G. Aging study: IGF-1 and body weight for 33 inbred strains of mice. MPD:Yuan1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Leduc,M.S., Hageman,R.S., Meng,Q., Verdugo,R.A., Tsaih,S.W., Churchill,G.A., Paigen,B. and Yuan,R. (2010) Identification of genetic determinants of IGF-1 levels and longevity among mouse inbred strains. *Aging Cell*, **9**, 823–836.
- Kang,H.M., Zaitlin,N.A., Wade,C.M., Kirby,A., Heckerman,D., Daly,M.J. and Eskin,E. (2008) Efficient control of population structure in model organism association mapping. *Genetics*, **178**, 1709–1723.
- Schadt,E.E. and Su,W.L. Whole liver gene expression in 22 inbred strains of mice. MPD:GX-Su1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Su,W.L., Sieberts,S.K., Kleinhanz,R.R., Lux,K., Millstein,J., Molony,C. and Schadt,E.E. (2010) Assessing the prospects of genome-wide association studies performed in inbred mice. *Mamm. Genome*, **21**, 143–152.
- Tabakoff,B., Bhave,S. and Saba,L. Whole brain gene expression in males of 20 inbred strains of mice. MPD:GX-Tabakoff1. *Mouse Phenome Database*. The Jackson Laboratory, Bar Harbor, ME, USA.
- Aylor,D.L., Valdar,W., Foulds-Mathes,W., Buus,R.J., Verdugo,R.A., Baric,R.S., Ferris,M.T., Frelinger,J.A., Heise,M., Frieman,M.B. *et al.* (2011) Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.*, **21**, 1213–1222.
- Philip,V.M., Sokoloff,G., Ackert-Bicknell,C.L., Striz,M., Branstetter,L., Beckmann,M.A., Spence,J.S., Jackson,B.L., Galloway,L.D., Barker,P. *et al.* (2011) Genetic analysis in the Collaborative Cross breeding population. *Genome Res.*, **21**, 1223–1238.
- Threadgill,D.W., Miller,D.R., Churchill,G.A. and de Villena,F.P. (2011) The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *Ilar J.*, **52**, 24–31.